Pestic. Sci. 1997, **49**, 29–34

Use of Graph-Theoretical Parameters to Predict Activity of Organosilane Insecticides

Gordon G. Cash

Environmental Effects Branch, Health & Environmental Review Division (7403),
US Environmental Protection Agency, 401 M Street, S.W., Washington, DC 20460, USA

(Received 7 January 1996; revised version received 14 June 1996; accepted 21 August 1996)

Abstract: Descriptions have appeared in the literature of insecticide analogues in which silicon was substituted for quaternary carbon. The interest in these compounds was that they appeared to be approximately as toxic to the target species as the carbon analogues but significantly less acutely toxic to fish. In the present paper, chemical graph theory and principal component analysis were applied to a series of these silane compounds to predict insecticidal activity, thereby unlocking the potential to identify promising candidates for chemical synthesis. Good correlations ($r > 0{\cdot}9$) were demonstrated, and possible physical interpretations of the relevant graph-theoretical parameters were discussed.

Key words: organosilane insecticide, graph theory, connectivity index

1 INTRODUCTION

In this paper, the author attempts to construct from molecular connectivity indices (MCI) equations predictive of the 50%-lethal concentrations (LC_{50}) against Mexican bean beetle (*Epilachna varivestis* Muls.) of structurally related pesticide analogues reported in the literature.[1,2] The actual variable correlated is not LC_{50} itself, but $\log(1/LC_{50})$. In this paper, the phrase 'dependent variable' refers to $\log(1/LC_{50})$.

Detailed rules for computing MCIs have been published elsewhere,[3,4] but two conventions are particularly relevant here. All MCIs ignore hydrogen atoms. Non-valence MCIs, ${}^{n}\chi_{x}$, treat all non-hydrogen atoms alike, recognizing only connections between atoms; valence MCIs, ${}^{n}\chi_{x}^{v}$, utilize different values for different elements. Thus, non-valence MCIs do not distinguish tetravalent silicon from tetravalent carbon (tetravalent in this context meaning bonded to four non-hydrogen atoms).

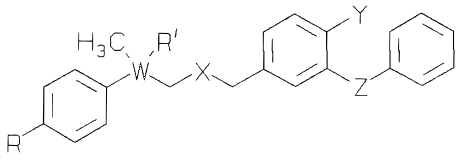
A large number of papers have been published reporting correlations of assorted molecular parameters with various experimental measures of biological activity. Many of these, however, suffer from a few recurring problems namely, a low ratio of data points to predictor variables, description of only a highly restricted set of compounds, use of parameters that are different to calculate, or use of parameters that bear no apparent physical relationship to interaction with biological systems. In the present study, the author reports good correlation with $\log(1/LC_{50})$ for a variety of dissimilar substituents while maintaining an acceptable ratio of predictors to data and utilizing computations well within the capability of a PC-class computer.

2 RESULTS AND DISCUSSION

Estimates of log P were calculated for the 18 compounds studied[5] (Table 1). Surprisingly, all attempts to find a correlation between log P (or $[\log P]^2$) and the dependent variable failed. Somewhat less surprisingly, there was no correlation between molecular weight and the dependent variable, either. A total of 38 MCIs were calculated for each compound, but six of these were zero because they only discerned rings of less than six members, of which there were none in the compounds studied here. That left 32 predictor variables for only 18 observations. The author pursued two avenues toward reducing the number of predictors. It is reasonable to question the appropriateness of generating so many potential predictor variables for so small a dataset. This is a common procedure with MCIs because these variables are often highly intercorrelated. In the present

Pestic. Sci. 0031-613X/97/$09.00

TABLE 1
Compounds Examined in this Study



Compound	*R*	*R'*	*W*	*X*	*Y*	*Z*
1	C_2H_5O	CH_3	Si	CH_2	H	CH_2
2	C_2H_5O	CH_3	Si	CH_2	H	CO
3	C_2H_5O	CH_3	Si	CH_2	H	NH
4	C_2H_5O	CH_3	Si	CH_2	H	NCHO
5	C_2H_5O	CH_3	Si	CH_2	H	NCH_3
6	C_2H_5O	CH_3	Si	CH_2	H	O
7	C_2H_5O	CH_3	Si	CH_2	F	O
8	CH_3O	CH_3	Si	CH_2	F	O
9	$(CH_3)_2CHO$	CH_3	Si	CH_2	F	O
10	H	CH_3	Si	CH_2	F	O
11	Cl	CH_3	Si	CH_2	F	O
12	CH_3	CH_3	Si	CH_2	F	O
13	CF_3	CH_3	Si	CH_2	F	O
14	C_2H_5O	C_2H_5	Si	CH_2	F	O
15	C_2H_5O	CH_2=CH	Si	CH_2	F	O
16	C_2H_5O	CH_3	C	O	H	O
17	C_2H_5O	CH_3	Si	O	H	O
18	C_2H_5O	CH_3	C	CH_2	F	O

study, it is very unlikely that the 32 MCIs mentioned above are all linearly independent. The expectation from the beginning is that many potential predictor variables will be generated, but only the best few will be chosen. As shown below, two variables relevant to the present dataset turned out to be the seventh-order chain MCIs, ${}^7\chi_{CH}$ and ${}^7\chi^v_{CH}$. If fewer MCIs had been generated in the beginning, these would have been missed altogether.

A commonly used means of reducing the dimensionality of problems such as this is principal component analysis (PCA). (Any textbook of multivariate statistics should offer an explanation of PCA and its use for reducing dimensionality.) Indeed, PCA in this case gave 16 meaningful principal components (PCs) and 16 for which the scores were all $0 \pm 0{\cdot}001$. Stepwise multiple regression with $F = 3$ to enter and to remove selected a subset of these which predicted $\log(1/LC_{50})$ very well ($r = 0{\cdot}9726$). Unfortunately, nine PCs were required, and there was no obvious point of diminishing return until the program stopped at nine. The author does not believe that an observation-to-predictor ratio of 2·0 is a meaningful result. Worse, only four of these PCs had eigenvalues >1. PCA runs with other sets of variables, however, are described below.

The second approach was a more *ad hoc* affair involving inspection of individual MCIs and their univariate statistics. Eight of the MCIs had very little variability, and none of the non-valence MCIs except ${}^7\chi_{CH}$ appeared very promising as a predictor. Therefore, additional work was done on a set of 13 MCIs consisting of ${}^7\chi_{CH}$, ${}^{0-6}\chi^v_P$, ${}^3\chi^v_C$, ${}^{4-6}\chi^v_{PC}$ and ${}^7\chi^v_{CH}$. Stepwise multiple regression on this set consistently identified ${}^7\chi_{CH}$, ${}^3\chi^v_P$, ${}^4\chi^v_{PC}$ and ${}^7\chi^v_{CH}$ as the important predictors. As discussed in more detail below, compounds 2 and 7 were already beginning to appear anomalous. Excluding compounds 2 and 7, eqn (1) gives the best result obtained with this set.

$$\log(1/LC_{50}) = 15{\cdot}479(3{\cdot}566)^7\chi_{CH} - 1{\cdot}039(0{\cdot}258)^3\chi^v_P + 1{\cdot}051(0{\cdot}295)^4\chi^v_{PC} - 25{\cdot}940(14{\cdot}499)^7\chi^v_{CH} - 0{\cdot}858$$
$$n = 16,\ s = 0{\cdot}308,\ r = 0{\cdot}942,\ F = 21{\cdot}83 \quad (1)$$

While an observation-to-predictor ratio of 4·0 is certainly better than one of 2·0, the author believes 4·0 is still too low. Principal component regression using this set of 13 independent variables and $F = 3$ to enter and to remove for the multiple regression step also gave an equation with four predictors which was clearly inferior to eqn (1).

Observing that the coefficients for ${}^3\chi^v_P$ and ${}^4\chi^v_{PC}$ in eqn (1) have opposite signs and about the same magnitude, one wonders whether $({}^4\chi^v_{PC} - {}^3\chi^v_P)$ would be an appropriate predictor variable, thus raising the observation-to-predictor ratio to 5·3 for $n = 16$. (The physical

meaning of $({}^4\chi^v_{PC} - {}^3\chi^v_P)$ is discussed below). Indeed, stepwise multiple regression with a set of 12 dependent variables like the set of 13 above, but including $({}^4\chi^v_{PC} - {}^3\chi^v_P)$ and excluding ${}^3\chi^v_P$ and ${}^4\chi^v_{PC}$ themselves, gave eqn (2).

$$\log(1/LC_{50}) = 15{\cdot}311(2{\cdot}672){}^7\chi_{CH} - 25{\cdot}012(7{\cdot}427){}^7\chi^v_{CH} + 1{\cdot}040(0{\cdot}247)({}^4\chi^v_{PC} - {}^3\chi^v_P) - 0{\cdot}839$$
$$n = 16,\ s = 0{\cdot}294,\ r = 0{\cdot}942,\ F = 31{\cdot}73 \quad (2)$$

It was especially promising that the standard error of the estimate was slightly lower in eqn (2) than in eqn (1).

Kier and Hall[3] discuss the utility of 'valence connectivity differential' variables (VCD), which they define as the difference between a non-valence MCI and its valence counterpart. In the light of their discussion, and despite the dissimilarity of the first two coefficients in eqn (2), $({}^7\chi_{CH} - {}^7\chi^v_{CH})$ was examined as an independent variable in place of its individual MCIs. The result of this exercise was eqn (3).

$$\log(1/LC_{50}) = 1{\cdot}260(0{\cdot}201)({}^4\chi^v_{PC} - {}^3\chi^v_P) + 15{\cdot}064(2{\cdot}774)({}^7\chi_{CH} - {}^7\chi^v_{CH}) - 1{\cdot}017$$
$$n = 16,\ s = 0{\cdot}306,\ r = 0{\cdot}932,\ F = 43{\cdot}056 \quad (3)$$

Compared to eqn (2), both s and r are slightly worse in eqn (3). Surprisingly, the 'adjusted R^2' value, which attempts to compensate for different observation-to-predictor ratios, is also worse (0·849 versus 0·860). Equation (3), of course, does have the higher ratio. The values are all so close that no convincing case can be made that either of eqns (2) and (3) is clearly the superior one. Their results are clearly different in one respect. With eqn (3), the residual for compound 18 is half as large again as any other residual. With eqn (2), there is no single observation that seems anomalous. Values for all the relevant MCI and VCD variables are listed in Table 2.

Repeating the regression analysis without compound 18 gave eqn (4).

$$\log(1/LC_{50}) = 1{\cdot}150(0{\cdot}158)({}^4\chi^v_{PC} - {}^3\chi^v_P) + 11{\cdot}683(2{\cdot}376)({}^7\chi_{CH} - {}^7\chi^v_{CH}) - 0{\cdot}7397$$
$$n = 15,\ s = 0{\cdot}235,\ r = 0{\cdot}935,\ F = 41{\cdot}675 \quad (4)$$

The adjusted R^2 value for eqn (4) is 0·853. Equations (2), (3) and (4) are presented together because the dataset is too small to support a clear judgement about which one is the best. The author's personal preference is eqn (4) because of the higher observation-to-predictor ratio and because, as discussed below, there is more reason to exclude compound 18 than just the large residual in eqn (3). A complete list of residuals from eqns (2), (3) and (4) is presented in Table 3.

It is appropriate to ask whether a variable like $({}^4\chi^v_{PC} - {}^3\chi^v_P)$ has any relevant physical meaning, or whether it might simply be fortuitously correlated with the dependent variable. Indeed, the author is unaware of any prior use of a composite variable of this type in structure–activity analysis. One needs to be particularly careful about fortuitous correlation with a dataset as

TABLE 2
Dependent Variable and Relevant MCI and VCD Variables

Compound	$log(1/LC_{50})^a$	${}^7\chi_{CH}$	${}^3\chi^v_P$	${}^4\chi^v_{PC}$	${}^7\chi^v_{CH}$	${}^4\chi^v_{PC} - {}^3\chi^v_P$	${}^7\chi_{CH} - {}^7\chi^v_{CH}$
1	−2·724	0·291	8·983	5·297	0·115	−3·686	0·176
2	−1·623	0·267	8·965	5·371	0·103	−3·594	0·164
3	−3·000	0·291	8·730	5·185	0·103	−3·545	0·188
4	−3·300	0·267	9·069	5·419	0·100	−3·650	0·167
5	−3·049	0·267	9·182	5·564	0·100	−3·618	0·167
6	−2·041	0·291	8·618	5·136	0·097	−3·482	0·194
7	−0·954	0·337	8·707	5·229	0·102	−3·478	0·235
8	−1·903	0·337	8·632	5·249	0·102	−3·383	0·235
9	−1·875	0·337	8·768	5·338	0·102	−3·430	0·235
10	−1·845	0·287	8·315	5·103	0·097	−3·212	0·190
11	−1·813	0·361	8·637	5·320	0·122	−3·317	0·239
12	−1·845	0·361	8·593	5·295	0·119	−3·298	0·242
13	−2·079	0·320	8·754	5·417	0·105	−3·337	0·215
14	−3·000	0·337	9·377	5·195	0·117	−4·182	0·220
15	−2·911	0·337	8·883	4·624	0·110	−4·259	0·227
16	−1·380	0·291	4·556	1·829	0·069	−2·727	0·222
17	−1·845	0·291	7·399	4·464	0·097	−2·935	0·194
18	−2·230	0·337	5·181	2·169	0·074	−3·012	0·263

[a] From References 1, 2.

TABLE 3
Measured and Predicted Values of $\log(1/LC_{50})$

Compound	*Measured*[a] $\log(1/LC_{50})$	*Equation (2)* $\log(1/LC_{50})$	*Diff.*[a]	*Equation (3)* $\log(1/LC_{50})$	*Diff.*[a]	*Equation (4)* $\log(1/LC_{50})$	*Diff.*[a]
1	−2·724	−3·093	0·369	−3·010	−0·286	−2·922	0·198
3	−3·000	−2·646	−0·354	−2·652	−0·348	−2·619	−0·381
4	−3·300	−3·048	−0·252	−3·100	−0·200	−2·985	−0·315
5	−3·049	−3·015	−0·034	−3·060	0·011	−2·949	−0·100
6	−2·041	−2·431	0·390	−2·482	0·441	−2·477	0·436
8	−1·903	−1·749	−0·154	−1·740	−0·163	−1·884	−0·019
9	−1·875	−1·797	−0·078	−1·799	−0·076	−1·938	0·063
10	−1·845	−2·211	0·366	−2·202	−0·357	−2·213	0·368
11	−1·813	−1·813	0·000	−1·596	−0·217	−1·761	−0·052
12	−1·845	−1·718	−0·127	−1·527	−0·318	−1·704	−0·140
13	−2·079	−2·036	−0·043	−1·983	−0·096	−2·065	−0·014
14	−3·000	−2·955	−0·046	−2·972	−0·028	−2·978	−0·022
15	−2·911	−2·860	−0·052	−2·964	0·053	−2·985	0·074
16	−1·380	−0·945	−0·435	−1·109	−0·271	−1·282	−0·098
17	−1·845	−1·862	0·017	−1·793	−0·052	−1·848	0·003
18	−0·230	−0·662	0·432	−0·850	0·620		

[a] References 1, 2.
[b] Experimental minus calculated.

small as this one. In the present case, however, $({}^4\chi^v_{PC} - {}^3\chi^v_P)$ does seem to reflect structural features one would expect to be relevant to biological activity. Contributors to ${}^4\chi^v_{PC}$ are five-atom (four-bond) groups connected in the manner of *iso*-pentane; contributors to ${}^3\chi^v_P$ are four-atom (three-bond) groups. The difference between them would reflect structural changes that increase the number of three-bond groups without correspondingly increasing the number of four-bond, *iso*-pentane-like groups. (The reverse situation is not physically possible). Because of the sign conventions used here, raising ${}^3\chi^v_P$ without raising ${}^4\chi^v_{PC}$ will cause eqn (4) to predict lower insect toxicity.

Examples from the present dataset can illustrate this situation. Substituting fluorine for hydrogen at Y (Table 1, compounds **6** and **7**) creates three new contributors to ${}^3\chi^v_P$, all containing fluorine, and three new contributors to ${}^4\chi^v_{PC}$, also all containing fluorine. Thus, this substitution is unlikely to have much effect on $({}^4\chi^v_{PC} - {}^3\chi^v_P)$. The same argument applies to substituting ethyl or vinyl for methyl at R′. On the other hand, substitutions at R or Z will have profoundly different effects on ${}^3\chi^v_P$ than on ${}^4\chi^v_{PC}$, particularly since these are valence indices and recognize heteroatoms as different from carbon. Therefore, these substitutions will effect changes in $({}^4\chi^v_{PC} - {}^3\chi^v_P)$. One can further argue that, since $({}^4\chi^v_{PC} - {}^3\chi^v_P)$ is well correlated with the dependent variable, substitutions at R and Z are the ones most likely to affect insect toxicity in compounds of this type.

Sieburth *et al.*[1,2] showed that increased insect toxicity was correlated with both increased steric bulk and electron-donating ability of the substituent at R. For the substituents under consideration here, increasing steric bulk at R is likely to have much the same effect on ${}^3\chi^v_P$ as on ${}^4\chi^v_{PC}$ and similarly on ${}^7\chi_{CH}$ as on ${}^7\chi^v_{CH}$. The entries in Table 2 support this conjecture. Thus, the correlations established in the present study neither support nor refute the steric bulk hypothesis set fourth by Sieburth *et al.* In general, substituting a heteroatom for carbon (while maintaining the same chain length) tends to decrease the valence MCIs while having no effect on the normal MCIs. From this argument, one might expect that substituting an electron-*withdrawing* group, such as F or Cl, at R would increase $({}^7\chi_{CH} - {}^7\chi^v_{CH})$, thus predicting higher toxicity. Indeed, substituting F for H at position Y clearly has this effect. The values of $({}^7\chi_{CH} - {}^7\chi^v_{CH})$ in Table 2, however, do not lend clear support for this effect at position R. Entries **10**–**13** are particularly relevant in this regard. Thus, the correlation with $({}^7\chi_{CH} - {}^7\chi^v_{CH})$ does not offer clear support or refutation for the electron-donating-substituent effect cited by Sieburth *et al.*,[1,2] either. Because ${}^3\chi^v_P$ and ${}^4\chi^v_{PC}$ are both valence MCIs, heteroatom substitution would have less effect on $({}^4\chi^v_{PC} - {}^4\chi^v_P)$.

The variable $({}^7\chi_{CH} - {}^7\chi^v_{CH})$ is an example of 'valence connectivity differential' as discussed by Kier and Hall.[3] Contributors to each of the two terms are connected groups of seven atoms that include at least one ring. In the present study, all such contributors are a benzene ring plus an atom bonded to it. Since the valence MCIs treat heteroatoms differently from carbon, and the non-valence MCIs do not, the largest differences in $({}^7\chi_{CH} - {}^7\chi^v_{CH})$ will be made by non-ether substitutions at R, carbon for silicon at W and fluorine for hydrogen at Y. In the first two cases, ${}^7\chi^v_{CH}$ changes while ${}^7\chi_{CH}$ does not; in the last case, ${}^7\chi_{CH}$ changes while ${}^7\chi^v_{CH}$ remains almost

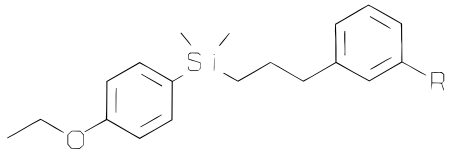
unchanged. Because of the relative values of the valence atom parameter δ^v for carbon, nitrogen and oxygen, substitution at Z will have less effect on $({}^7\chi_{CH} - {}^7\chi^v_{CH})$. In retrospect, the author believes that, for this set of compounds, $({}^4\chi^v_{PC} - {}^3\chi^v_P)$ and $({}^7\chi_{CH} - {}^7\chi^v_{CH})$ are likely to reflect structural changes important in determining biological activity.

In analyzing other datasets, the author had some success with transforming both dependent and independent variables so as to minimize the Kolmogorov–Smirnov score for normality.[6,7] These transformations took the form $(X + a)$,[b] where $a > 0$ only if some X were negative. With the present dataset, however, the dependent variable was already near its minimum Kolmogorov–Smirnov score, and the independent variables generally approached their minimum scores as b approached zero. Nevertheless, with regard to the Kolmogorov–Smirnov score, it is noteworthy that, of all the non-principal-component variables studied, only ${}^0\chi^v_P$, $({}^4\chi^v_{PC} - {}^3\chi^v_P)$, $({}^7\chi_{CH} - {}^7\chi^v_{CH})$ and $\log(1/LC_{50})$ pass the Kolmogorov–Smirnov test for normality. For the present dataset, $({}^4\chi^v_{PC} - x^3\chi^v_P)$ and $({}^7\chi_{CH} - x^7\chi^v_{CH})$ were also examined as variables, where x in each case was the ratio of the means of the individual terms, thus effectively putting the two terms on the same scale. Both of these variables were less useful predictors of insect toxicity than were their non-rescaled counterparts.

As mentioned above, in calculating a bivariate correlation with $({}^4\chi^v_{PC} - {}^3\chi^v_P)$ and $({}^7\chi_{CH} - {}^7\chi^v_{CH})$, compounds **2**, **7** and **18** fit much less well than the remaining 15. Below, some observations are made about these three compounds which the author believes justify regarding them as outliers. There is no conspicuous structural element that distinguishes compounds **7** and **18**, but inspection of Table 1 reveals that the two compounds are structurally identical except for silicon versus carbon at position W. Furthermore, calculating the dependent variable for the two using eqn (4) gives answers that are both in error by about the same amount (+1·039 and +0·900 log unit). This observation is consistent with the notion that whatever the mechanistic reason for the observed increase in insect toxicity, it operates for both compounds to about the same extent, as one would expect for two structures so nearly identical. The author therefore speculates that some unknown mechanism enhances the biological activity of these two particular structures.

Compound 2, unlike all the others, is a substituted benzophenone. One therefore wonders whether enzymatic cleavage of the ketone might give rise to the anomalous (as defined by eqn (4)) insect toxicity. To explore this speculation further, the dependent variable was calculated according to eqn (1) for five possible cleavage products and the results are presented in Table 4. Indeed, oxidative cleavage of the benzophenone to the substituted phenol would give a calculated result only 0·324 log unit different from the observed value, a

TABLE 4
Possible Products from Cleavage of Compound 2



R	*Calculated* $\log(1/LC_{50})$	*Residual*	*RSE*[a]
H	−2·663	1·040	4·43
OH	−1·947	0·324	1·38
CHO	−2·415	0·792	3·37
CH_2OH	−2·514	0·896	3·82
COOH	−2·527	0·904	3·85

[a] Relative standard error = residual ÷ standard error of the estimate from eqn (1).

differences well within the range of the other residuals. A search of the literature revealed no precedent for enzymatic oxidative cleavage of diaryl ketones to phenols, so any suggestion that such a reaction occurs in *E. varivestis* would be entirely speculative. While this cleavage may or may not be a metabolic pathway here, the author believes that the use of observed MCI correlations to suggest possible metabolic transformations is valid.

Given Kier and Hall's discussion of VCDs and the relative success of $({}^7\chi_{CH} - {}^7\chi^v_{CH})$ as a predictor, the other possible VCDs were also investigated. For each MCI from the set of ${}^{0-6}\chi_P$, ${}^3\chi_C$, ${}^{4-6}\chi_{PC}$ and ${}^7\chi_{CH}$, the valence counterpart was subtracted to give 12 VCD variables. To this set was also added $({}^4\chi^v_{PC} - {}^3\chi^v_P)$. Stepwise multiple regression with $F = 3$ to enter and to remove, and excluding compounds **2** and **7**, gave eqn (3) or (4) depending on whether compound **18** was included. Thus, none of the new VCDs predicted $\log(1/LC_{50})$ as well as those already discovered. The same operation with all 18 compounds selected the same two independent variables and no others. In that run, compounds **2**, **7** and **18**, in that order, had the largest residuals. For the sake of completeness, the result of that run is presented as eqn (5).

$$\log(1/LC_{50}) = 1{\cdot}212(0{\cdot}326)({}^4\chi^v_{PC} - {}^3\chi^v_P) + 12{\cdot}398(4{\cdot}398(4{\cdot}095)({}^7\chi_{CH} - {}^7\chi^v_{CH}) - 0{\cdot}493$$
$$n = 18,\ s = 0{\cdot}498,\ r = 0{\cdot}812,\ F = 14{\cdot}573 \quad (5)$$

Principal component regression on this set of 13 independent variables produced eqn (6), the correlation coefficient and standard error of the estimate for which are nearly identical to those of eqn (3). Compounds **2** and **7** were excluded from the multiple regression step but not from the PCA step. As in eqn (3), the residual for compound **18** was not notably larger than some of

the others. In fact, the plots of residuals for eqns (3) and (6) look a great deal alike.

$$\log(1/LC_{50}) = 0{\cdot}155(0{\cdot}023)PC1 + 0{\cdot}597(0{\cdot}090)PC3 - 2{\cdot}211$$
$$n = 16,\ s = 0{\cdot}302,\ r = 0{\cdot}934,\ F = 44{\cdot}290 \quad (6)$$

That eqns (3) and (6) are so similar is a surprising result because, while $({}^4\chi^v_{PC} - {}^3\chi^v_P)$ and $({}^7\chi_{CH} - {}^7\chi^v_{CH})$ are the most important contributors to the third principal component (PC3), they are the least important contributors to PC1. The other 11 variables contribute about equally to PC1. Since the respective eigenvalues are 10·64 and 0·77, eqn (6) seems to say that the other 11 VCDs are collectively more important than $({}^4\chi^v_{PC} - {}^3\chi^v_P)$ and $({}^7\chi_{CH} - {}^7\chi^v_{CH})$ and PCA allows this fact to appear in the regression equation without increasing the number of independent variables.

The author admits lacking a credible structural explanation for why eqns (2), (3) and (6) predict the activity of compound **18** as well as they do. It simply makes no sense that equations derived to predict the activities of silicon-containing materials would succeed for the carbon-containing compound **18** while failing for the silicon-containing but otherwise identical compound **7**. In a dataset of this small size, one is greatly tempted to attribute a finding such as this to coincidence.

Examination of plots of $\log(1/LC_{50})$ versus one or two other variables uncovered a very good correlation ($r = 0{\cdot}955$) between $\log(1/LC_{50})$ and $({}^4\chi^v_{PC} - {}^3\chi^v_P)$ for an $n = 11$ set that excluded compounds **1**, **2**, **3**, **4**, **5**, **7** and **18**. As discussed above, compounds **7** and **18** are also anomalous in other spaces. The remaining five compounds are just those that have Z other than oxygen. This correlation is evidence that $({}^4\chi^v_{PC} - {}^3\chi^v_P)$ alone adequately describes all the structural changes within this dataset except the ones at Z. This statement immediately begs the question whether $({}^7\chi_{CH} - {}^7\chi^v_{CH})$ alone correlates well with compounds **1** through **5**, but that correlation coefficient is $< 0{\cdot}5$, either with or without compound **2**.

3 CONCLUSIONS

Graph-theoretical parameters and principal components derived from them showed good correlation ($r > 0{\cdot}9$) with biological activity for a group of silicon-containing pesticides. A composite parameter of a type not previously used in structure–activity analysis, namely $({}^4\chi^v_{PC} - {}^3\chi^v_P)$, is shown to be a good predictor of biological activity for this group and to describe relevant structural features. Thus, these parameters may be useful in predicting activity for other compounds in the same class that are candidates for synthesis. Also based on these correlations, the author postulates a possible metabolic transformation of one compound prior to its entry into the pesticidal metabolic pathway common to the group.

REFERENCES

1. Sieburth, S. McN., Manly, C. J. & Gammon, D. W., Organosilane insecticides. Part I: Biological and physical effects of isosteric replacement of silicon for carbon in Etofenprox and MTI-800. *Pestic Sci.*, **28** (1990) 289–307.
2. Sieburth, S. McN., Langevine, C. N. & Dardaris, D. M., Organosilane insecticides. Part II: Chemistry and structure–activity relationships. *Pestic Sci.*, **28** (1990) 309–19.
3. Kier, L. B. & Hall, L. H., *Molecular Connectivity in Chemistry and Drug Research*. Academic Press, New York, NY, 1976.
4. Kubinyi, H., Lipophilicity and drug activity. *Progr. Drug Res.*, **23** (1979) 92–198.
5. Meylan, W. M. & Howard, P. H., Atom/fragment contribution method for estimating octanol-water partition coefficients. *J. Pharm. Sci.*, **84** (1995) 83–92.
6. Cash, G. G. & Breen, J. J., Correlation of graph-theoretical parameters with biological activity. *J. Chem. Inf. Comput. Sci.*, **33** (1993) 275–9.
7. Cash, G. G., Correlations of molecular connectivity indices with toxicities of organic nitriles. In *Biomarkers of Human Exposure to Pesticides, ACS Symposium Series 542*, ed. M. A. Saleh, J. N. Blancato & C. H. Nauman. ACS Books, Washington, DC, 1994, pp. 65–75.